C:\sharad\CV\AWS\_CV\Capgemini\AWS\_CLI\_work\github\_work\github-to-aws\.github\workflows

Create directory github-to-aws.

Create directory \.github\workflows (Note we need .github and then workflow)

Create ci-cd.yml :-

* When push or pull is done script will kick
* Cfn -lint will check the script “crawler\_buildspec\_S3.yml” is as per standard i.e. no syntax issue , no indentation issue etc.
* Set AWS\_ACCESS\_KEY\_ID and AWS\_SECRET\_ACCESS\_KEY in Github and use it in “Deploy CloudFormation stack” command where we are executing script “crawler\_buildspec\_S3.yml” **template-file .github/workflows/crawler\_buildspec\_S3.yml**

Create crawler\_buildspec\_S3.yml :-

This is the yaml script which creates AWS resources in AWS and will be called from ci-cd.yml

We can create all the resource form CLoudFormaiton and create Stack but this will be manually.

Create Script my-pyspark-script.py :-

This script is to pick the table information in AWS Glue , created by Crawler and will do transformation on data.csv file and final result goes to new S3 bucket “my-example-s3-bucket-2025-output”.

This script is called from **crawler\_buildspec\_S3.yml ( once all resources are created).**

Create script my-pipeline-stack.yml :-

This script is to run create pipeline. Which will run **crawler\_buildspec\_S3.yml.**

**All these codes are in Github :-**

1. **ci-cd.yml**
2. **my-pipeline-stack.yml**
3. **crawler\_buildspec\_S3.yml**

**set token in Github which will be used in pipeine command** :-

Goto STThukral logo on right -> Settings -> Developer settings -> Personal Access Tokens -> Tokens Classic

Check below in check box and create token (bold one are important)

*admin:enterprise,****admin:org****, admin:org\_hook,****admin:public\_key****, admin:repo\_hook,****repo****, user, workflow, write:discussion, write:packages*

**Goto command prompt and set directory :-**

C:\Users\yashi>cd C:/sharad/CV/AWS\_CV/Capgemini/AWS\_CLI\_work/github\_work/github-to-aws/.github/workflows

C:\sharad\CV\AWS\_CV\Capgemini\AWS\_CLI\_work\github\_work\github-to-aws\.github\workflows>code .

This will open all the files in workflows in Visual code.

git add crawler\_buildspec\_S3.yml

git commit -m "Adding crwaler buildspec"

git push origin main

git add my-pipeline-stack.yml

git commit -m "Add CI/CD pipeline"

git push origin main

Bold one is the Github token

$ aws cloudformation create-stack --stack-name my-pipeline-stack --template-body file://my-pipeline-stack.yml --parameters ParameterKey=GitHubToken,ParameterValue="**ghp\_ldESBXYJ9J0lupsXXg11Ds9azYVwS10eeUAR**" --capabilities CAPABILITY\_NAMED\_IAM

This will create pipeline you can check from Console check from CODE BUILD, CODE DEPLOY and CODE PIPELINE

First my-pipeline-stack is created , once you do any minor changes in **crawler\_buildspec\_S3.yml** and pipeline will run and will create another stack **my-cloudformaiton-stack**

A computer screen with text

AI-generated content may be incorrect.

Bucket  [my-example-s3-bucket-2025-output](https://eu-west-2.console.aws.amazon.com/s3/buckets/my-example-s3-bucket-2025-output?region=eu-west-2&bucketType=general)

And [my-example-s3-bucket-2025](https://eu-west-2.console.aws.amazon.com/s3/buckets/my-example-s3-bucket-2025?region=eu-west-2&bucketType=general) created under stack process

S3 bucket [s3-lambda-function-code-st](https://eu-west-2.console.aws.amazon.com/s3/buckets/s3-lambda-function-code-st?region=eu-west-2&bucketType=general) created manually containing code :-

[create\_crawler.zip](https://eu-west-2.console.aws.amazon.com/s3/object/s3-lambda-function-code-st?region=eu-west-2&bucketType=general&prefix=create_crawler.zip) and [my-pyspark-script.py](https://eu-west-2.console.aws.amazon.com/s3/object/s3-lambda-function-code-st?region=eu-west-2&bucketType=general&prefix=my-pyspark-script.py)

A screenshot of a computer

AI-generated content may be incorrect.

Github Actions showing all went fine :-

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

Upload Data.csv file in S3 bucket [my-example-s3-bucket-2025](https://eu-west-2.console.aws.amazon.com/s3/buckets/my-example-s3-bucket-2025?region=eu-west-2&bucketType=general) . By doing this Crawler will run and Table and Database will be created in AWS Glue.

Running Script manually my-pyspark-script.py running manually from AWS as need to check as not running automatically form Lambda.

Goto AWS Glue -> ETL jobs -> select [MyPySparkETLJob](https://eu-west-2.console.aws.amazon.com/gluestudio/home?region=eu-west-2#/editor/job/MyPySparkETLJob) and “Run job”

A screenshot of a computer

AI-generated content may be incorrect.



Output :-

A screenshot of a computer

AI-generated content may be incorrect.

Visual Code screen for reference :-

A screenshot of a computer program

AI-generated content may be incorrect.

Github reference :-

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.